Title of disclosure (in English)
Method of Self Enhancement of Search Result by Analyzing the System Log

1. Describe your invention, stating the problem solved (if appropriate), and indicating the advantages of using the invention.
Modern document search systems allow to perform full text search on collections of documents and return the documents that contain the search query terms. The relevancy of the search results depends on many factors, in particular, on the specificity of the search query. If the search query was specific enough, the probability of getting relevant results in the first page would be higher. For example, the probability of getting documents on 'Java exception handling' in the first page is higher for the query that contains 'Java exception' than for the query that contains only 'exception'. At the same time, some relevant documents may not be returned in response to a specific search query, because they do not contain certain combination of terms, or describe the same topic in different words. For example, if the query was 'video player for PC', the search engine will not be able to find and return relevant documents that contain terms like 'DVD driver' or 'multimedia software'. Figure 1 illustrates this example, showing that some relevant documents may not be returned to the user.
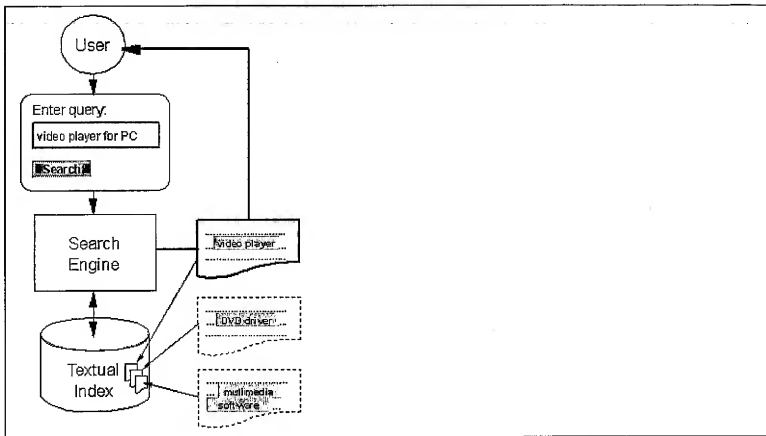

The proposed method allows to improve automatically the probability of returning relevant search results in response to a specific search query. This improvement is achieved by analyzing the log of the document search system, identifying and enhancing user search queries with existing techniques, such as glosary terms and synonyms, classifying augmented queries, locating relevant documents that were not returned by search system, and enhancing the metadata of the related documents in the search index to ensure that they will be returned next time in response to a similar query. The mechanism of search index/meta data self enhancement may become a part of autonomic document search system, providing a systematic way of improving user search experience.

2. How does the invention solve the problem or achieve an advantage,(a description of "the invention", including figures inline as appropriate)?
The automatic search index/meta data self-enhancement system consists of the following modules:
- search system log analyzer, which periodically looks through the search system log, and identifies search queries that did not bring satisfactory results.

- search query analyzer, this module applies known query enhancement techniques by using glosary terms, synonyms, known typos, translated words, etc. Once the queries are enhanced, they are automatically categorized and assigned to one or more subject areas.
- relevant document finder, based on the enhanced queries and their categorization, documents that were not previously found are detected and they are flagged for processing, which links the document to the query terms in the search index.
- search index / meta data enhancer, this module will enhace the metadata of the documents based on the enhanced query terms and the Search Index is updated to reflect this new keywords which will allow for the documents to be returned when similar searches are entered.
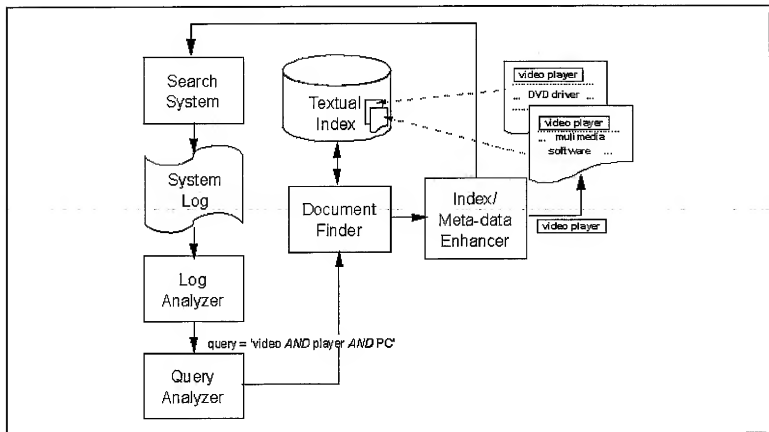
Novelty:
- Using this method to directly improve effectiveness of the search index by considering user queries which are enhanced and categorized and then relate this queries to specific documents.
-
- Main differentiators of this disclosure are:
- analysis is based on real customer query logs, so the enhancement directly results in customer satisfaction improvement;
- query log analysis and identification of relevant documents are performed off-line, so the response time is not affected;
- search system can apply this method in an autonomous way, which means that it is to become an important of an Autonomic Knowledge System that learns from user interactions and is based on detection of unsatisfactory results.
-

3. If the same advantage or problem has been identified by others (inside/outside IBM), how have those others solved it and does your solution differ and why is it better?
The problem is known but there is no autonomic solution that allows the system system to handle this problem by itself.

4. If the invention is implemented in a product or prototype, include technical details, purpose, disclosure details to others and the date of that implementation.
dBlue is GA'd on ▉▉▉▉. The improved version will be available ▉▉▉